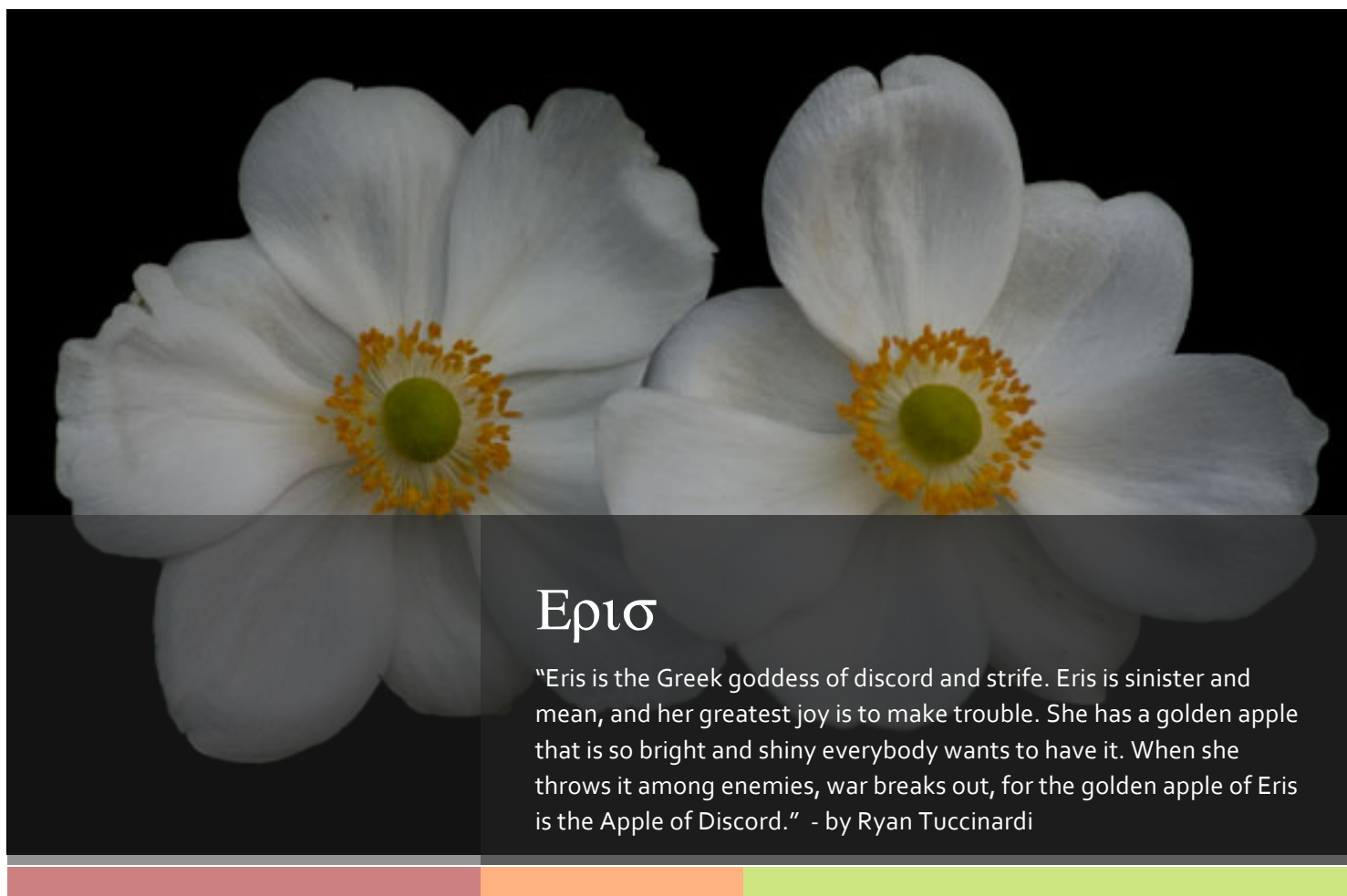


Eris

V 1.0



Eris



Eris

"Eris is the Greek goddess of discord and strife. Eris is sinister and mean, and her greatest joy is to make trouble. She has a golden apple that is so bright and shiny everybody wants to have it. When she throws it among enemies, war breaks out, for the golden apple of Eris is the Apple of Discord." - by Ryan Tuccinardi

Eris, which takes the name of Greek goddess of discord, is protein stability prediction and design software. Eris calculates the change of the protein stability induced by mutations (Delta-Delta-G, $\Delta\Delta G$) utilizing the recently developed Medusa modeling suite. In our test study, the $\Delta\Delta G$ values of a large dataset (>500) were calculated and compared with the experimental data and significant correlations are found. The overall correlation coefficient was 0.75. Compared with many other stability prediction methods, Eris is not specifically trained using experimental protein stability data and should be valid for a wider range of proteins. Furthermore, Eris models backbone flexibility, which turns out to be crucial for $\Delta\Delta G$ estimation of small-to-large mutations.

More details are available in the following references:

S. Yin, F. Ding, and N. V. Dokholyan, "Eris: an automated estimator of protein stability" *Nature Methods* **4**, 466–467 (2007)

S. Yin, F. Ding, and N. V. Dokholyan, "Modeling backbone flexibility improves protein stability estimation" *Structure*, **15**: 1567–1576 (2007)



Medusa force field

Eris uses a united atom model and expresses the free energy of the protein as a weighted sum of van der Waals (VDW) interaction, solvation energy, hydrogen bonding, and backbone-dependent statistical energy for any given amino acid and rotamer state. The unfolded state is assumed to be unstructured and fully-solvated and, thus, the reference energy of the unfolded proteins is the linear sum over all the amino acids, which is also parameterized in the Medusa force field. The weight coefficients are trained on 34 high-resolution X-ray protein structures so that occurrences of native amino acids sequences are maximized for all proteins.

Fixed-backbone method

In a fixed backbone calculation, Eris starts from the native structure and mutate the native residues. After the substitution, the side chain dihedrals of all residues are minimized using a Monte Carlo (MC) simulated annealing procedure. To search the dihedral space faster and more effectively, all the rotamer angles are divided into discrete rotamer states using the backbone-dependent rotamer library. Due to the stochastic nature of the method, the simulation is repeated 20 times from the randomly assigned initial side chain rotamers and the final free energy is taken to be the average over these 20 runs. The $\Delta\Delta G$ value is calculated as the difference between the ΔG of the mutant and the wild type as

$$\Delta\Delta G = \Delta G(\text{mutant}) - \Delta G(\text{wild type}).$$

To reduce the sampling space, only residues within a cutoff distance (10 Angstrom) are allowed to move during MC steps. The distance is measured as the smallest distance between any pair of atoms between the two residues.

Flexible-backbone method

In flexible-backbone stability prediction, Eris performs similar calculations as that in the fixed-backbone approach, except that it

allows the backbone dihedral angles to be relaxed if the acceptance rate during the MC loop is below a pre-defined threshold. The threshold used in Eris protocol is 0.05; such a low acceptance rate usually indicates strains on the backbone by the mutated residue. The relaxation is done by a conjugate-gradient minimization over all backbone dihedral angles at the end of the MC loop.

Pre-relaxation

The backbone pre-relaxation is implemented by performing the flexible-backbone calculation on the wide-type protein structure for 20 times. The redesigned structure with the lowest free energy is selected as the pre-relaxed structure.

Single-mutation Scanning

Eris calculates the $\Delta\Delta G$ of all single mutations throughout the whole protein sequence or on the specified sites. The aim is often to find stabilizing mutations in a relatively easy and fast way. To increase the speed, Eris uses only the fixed backbone method and the simulation is performed only once. Eris also uses a 10 Angstrom cutoff distance to reduce the number of movable residues.

Due to the limitations of the fixed-backbone method and sampling inefficiency, atomic clashes may not be resolved during structure minimization. In this case, the estimate of $\Delta\Delta G$ is not reliable. These clashes can be identified by checking the van der Waals repulsion (VDWR) energy terms in the results. Eris uses a cutoff of 10 kcal/mol, i.e., any mutation that results in a VDWR energy change of more than 10 kcal/mol is considered unreliable.

After excluding the unresolved clashes, the predicted stabilizing mutations will be listed in the output directory in `eris_working_directory/jobid`. Note that the prediction uncertainty can be large and checking the details of the energetic contributions will help assess the results. It is also observed that the predictions



are relatively more accurate for buried residues than exposed residues.

Protein Design

In the protein design mode, the Eris program attempts to find the optimal amino acid sequences for the given protein backbone. The search can be done with either fixed backbone (C, O, CA and N positions fixed during design), or flexible backbone (allowing small adjustments of the backbone atoms to minimize energy) protocol. The allowed sequences are listed in a file called "design table". The explanation of the design table is described next.

Format of the design table files:

```
DEFAULT KEYWORD [entry]
INDEX KEYWORD [entry]
INDEX KEYWORD [entry]
...
```

Valid KEYWORDS are:

ALLAA: all available amino acids and the corresponding rotamers
 NATAA: fixed with native amino acid, but with all its available rotamers
 NAROT: fix the amino acid in its native rotamer but with sub-rotamer allowed
 FIXNR: fix the amino acid in its native rotamer without sub-rotamer motion
 POLAR: polar amino acids and their rotamers
 HYDPH: hydrophobic amino acids
 AROMA: aromatic amino acids
 PIKAA: with selected amino acids

For PIKAA, selected amino acids are followed with a string entry using single letter representation.

For POLAR residues, we include SER, THR, GLN, GLU, ASN, ASP, LYS, ARG, HIS.

For HYDPH residues, we include GLY, ALA, MET, VAL, LEU, ILE, PHE, TYR, TRP, PRO, CYS.

FOR AROMA residues, we include PHE, TYR, TRP, HIS.

INDEX can be a single number (m), a set of number separated by comma (m,n,..), or a range defined as m-n (m<n); OR, a mixture such as (a-b,c,d,e-f).

"DEFAULT" can be used to represent all residues that have not been explicitly specified.

----- Design table example 1 -----

```
1-2 NATAA
3 ALLAA
4 POLAR
5 HYDPH
6 PIKAA CM
7 NAROT
8 AROMA
9 PIKAA QEND
```

----- Design table example 2 -----

```
DEFAULT NATAA
3 PIKAA QEND
```

Output:

The protein structures and detailed energy output will be stored in ./ddg_working_dir/jobid directory, where "jobid" is a unique name generated by the program or user-specified for distinguishing different $\Delta\Delta G$ estimation jobs.

Method modifiers:

-f, --flexbb flexible backbone.

Modes of Operation

“But when Eris leaves, the state of harmony that was previously enjoyed returns, for harmony has no part in discord.”

- Unknown

Eris v 1.0 uses the "eris_working" directory (Users can specify a different directory using "-w" option) to perform the specified calculations. If the directory does not exist, a new one will be created. For each task, eris-1.0 will assign a unique "jobid" (it can be specified by using "-j" option). The related output files will be stored in the subdirectory of "eris_working/jobid/". The file "eris_working/jobid/job" keeps the record of all performed simulations.

During Eris simulations, the program prints out the log to the standard output. At the end of the simulation, a brief description of the simulation type and the result is appended to a summary file, ".eris_results.txt". The output includes the jobid, job type, input pdb file, and brief summary of results.

The mutant protein structures for each run of repacking and detailed energy output will be stored in ".eris_working/jobid" directory. Next, the meaning of each mode of Eris' operation is explained.

ddg

For specified mutations (single or multiple substitutions) from the command line argument, eris-1.0 repacks the sidechain 20 times using simulated annealing and computes stabilities by averaging all the conformations for both wild-type ("eris_working/jobid/nat/nat.run00[00-19]") and mutant ("eris_working/jobid/mut/mut.run00[00-19]"). The stability change, $\Delta\Delta G$, is computed as the difference between the average stabilities of mutant and wild type. Each of these ".run" files contains the atomic conformation in PDB format.

"eris_working/jobid/mutation.dt" is the design table (see Method section) used for repacking calculations of the mutant protein.

"eris_working/jobid/native.dt" is the design table used for repacking calculations of the wild-type protein.

The entries in "eris_results.txt" file summarizes the results from the $\Delta\Delta G$ calculation. For example,

Modes of Operation

000002 ddg 1cka-ren.pdb V54I 0.699

0.699 -2.480 1.978 0.516 0.000 -0.287 0.272 0.492 0.473 0.585 -0.145

The job (jobid=000002) is the ddg calculation for input pdb of 1cka-ren.pdb. The mutation (V54I) is the substitution of Valine at position 54 to Isoleucine. The total ddg equals to 0.699, i.e. slightly destabilizing.

The second line corresponds to energy decomposition of total ddG. From the left to the right, the 11 values correspond to total $\Delta\Delta G$, and its decomposition of vdw attraction, vdw repulsion, solvation, backbone hydrogen bonds, backbone-sidechain hydrogen bonds, sidechain-sidechain hydrogen bonds, backbone-dependent statistical energy for amino acid, backbone-dependent statistical energy of the rotamer, reference energy of the unfolded states, and the correction for reference energy, respectively. For the meaning of each energy terms, please refer to:

F. Ding and N. V. Dokholyan, "Emergence of protein fold families through rational design" *Public Library of Science Computational Biology*, **2**: e85 (2006)

The details of the correction of reference energies can be found in:

S. Yin, F. Ding, and N. V. Dokholyan, "Modeling backbone flexibility improves protein stability estimation" *Structure*, **15**: 1567-1576 (2007)

scan

The scan action of eris-1.0 searches for the stabilizing single mutations with the protein backbone fixed. The scan mode of eris does not support flexible-backbone protocol. The positions to be scanned can be given from the command line (-s). If positions are not explicitly specified, all the residues are scanned.

The entry in the "eris_results.txt" file summarizes the scan simulations and points to "eris_working/jobid/output/ddgStabilizing.dat" for

further details.

For each position, the native amino acids are substituted to all other 19 types of amino acids. The atomic structures of repacked conformations are stored in "eris_working/jobid/output". The stabilizing mutations are listed in "eris_working/jobid/output/ddgStabilizing.dat". Each line in the ddgStabilizing.dat specifies the stabilizing single mutations. In the same line, the numbers starting from the second column correspond to the total stability change and its decomposition of vdw attraction, vdw repulsion, solvation, backbone hydrogen bonds, backbone-sidechain hydrogen bonds, sidechain-sidechain hydrogen bonds, backbone-dependent statistical energy for amino acid, backbone-dependent statistical energy of the rotamer and the reference energy of the unfolded states, respectively.

design

The design action of Eris searches the lowest-energy sequence that satisfies the constraints listed in the design table (-d). The details of the formation/syntax of the design table are given in the "protein design" subsection of Methods in this Manual. Typing "eris design -h" will also bring the corresponding details of design table.

The results of 20 rounds of fixed-backbone or flexible-backbone redesign simulations are given as "eris_working/jobid/design/design.run00[00-19]". In each ".run" file, the first 20 lines record the temperature, Monte Carlo acceptance rate, total energy and its decompositions at each of the 20 simulated annealing steps. For example, the 20th line gives the final energy and its decompositions. The following command will give the total energy and its decomposition of the a given ".run" file (design.run0000):

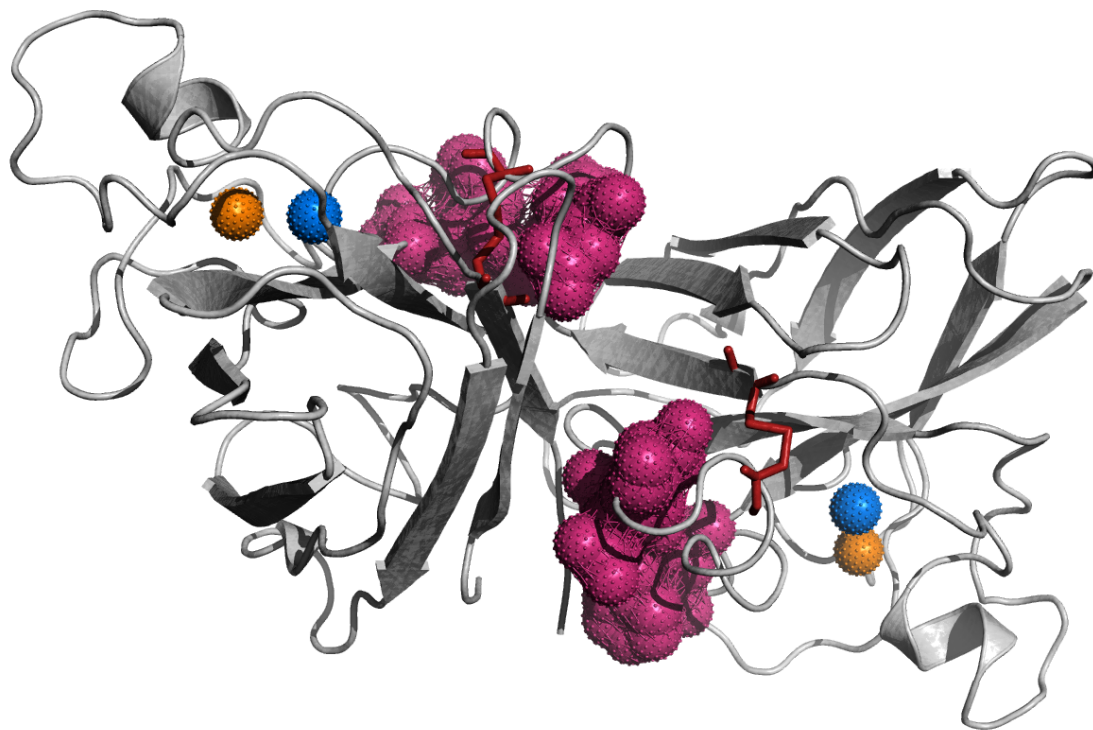
```
grep -m 1 "^19 " design.run0000 | cut -d" " -f4-
```

Modes of Operation

The output will look as follows:

```
-186.609 -260.056 7.98008 121.42 -68.3108 -10.6787 -8.85888 41.6438  
14.573 -24.3212
```

The terms correspond to the total stability, vdw attraction, vdw repulsion, solvation, backbone hydrogen bonds, backbone-sidechain hydrogen bonds, sidechain-sidechain hydrogen bonds, backbone-dependent statistical energy for amino acid, backbone-dependent statistical energy for the rotamer and reference energy of the unfolded state, respectively.



Molecules in Action, LLC



Molecules in Action, LLC

1530 Mill Valley Road
Chapel Hill, NC 27516

<http://moleculesinaction.com>

